Semantic Data

Introduction

Jean-Louis Binot

Agenda

1 Introduction to semantic data

2 Outline of the course

What is common in these situations (1) ?

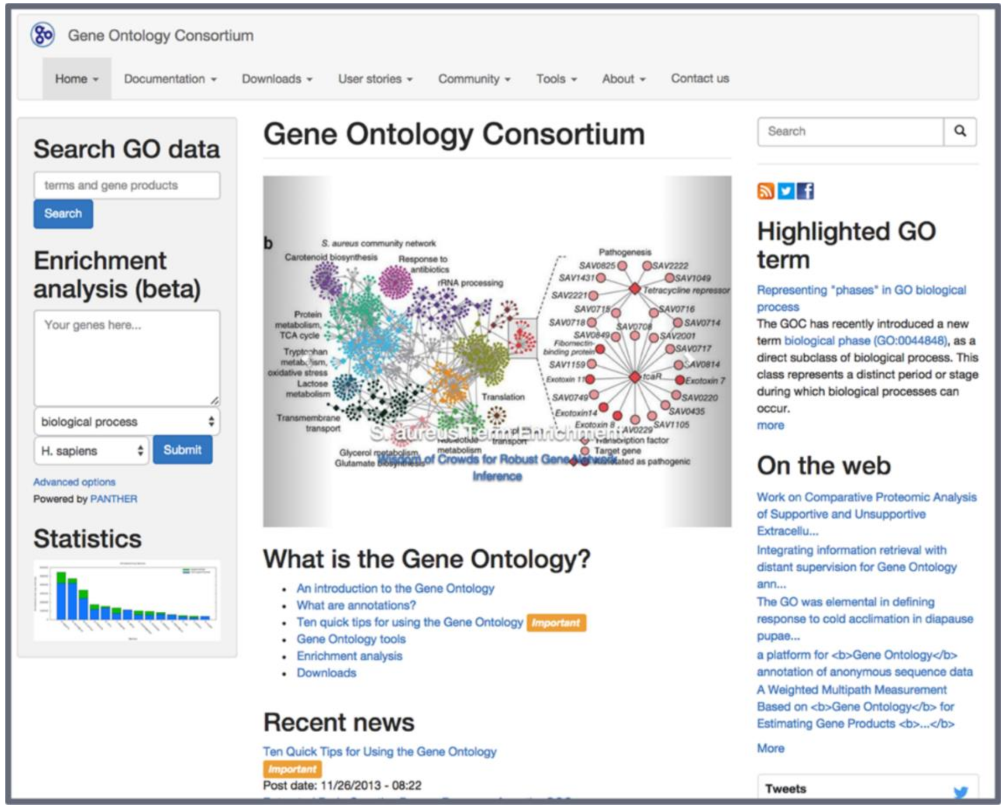
Building Watson: An Overview of the DeepQA Project.

“The overarching principles in DeepQA are massive parallelism, many experts, pervasive confidence estimation, and integration of shallow and deep knowledge : … balancing the use of strict and shallow semantics, leveraging many loosely formed ontologies …

such as DBpedia, WordNet, and the Yago ontology.”

(Ferruci et al., AI Magazine, 2010)

What is common in these situations (2) ?



Search GO data

Enrichment analysis (beta)

Advanced options
Powered by PANTHER

Statistics

Gene Ontology Consortium

Wisdom of Crowds for Robust Gene Network Inference

What is the Gene Ontology?

- An introduction to the Gene Ontology
- What are annotations?
- Ten quick tips for using the Gene Ontology Important
- Gene Ontology tools
- Enrichment analysis
- Downloads

Recent news

Ten Quick Tips for Using the Gene Ontology
Important
Post date: 11/26/2013 - 08:22

Highlighted GO term

Representing "phases" in GO biological process
The GOC has recently introduced a new term biological phase (GO:0044848), as a direct subclass of biological process. This class represents a distinct period or stage during which biological processes can occur.
more

On the web

Work on Comparative Proteomic Analysis of Supportive and Unsupportive Extracellu...
Integrating information retrieval with distant supervision for Gene Ontology ann...
The GO was elemental in defining response to cold acclimation in diapause pupae...
a platform for <b>Gene Ontology</b> annotation of anonymous sequence data
A Weighted Multipath Measurement Based on <b>Gene Ontology</b> for Estimating Gene Products <b>...</b>
More

(Gene Ontology consortium : going forward, Nucleic acids research 2014*)*

The Gene Ontology project provides controlled vocabularies of defined terms for a consistent description of gene products across databases.

"The Gene Ontology (GO) has become the most cited ontology in the PubMed/MEDLINE database." (Bodenreider, 2008).

"The Web Ontology Language (OWL) is a crucial component of the GO internal infrastructure."

(Gene Ontology Annotations and Resources, Nucleic Acids Research 2012)

Ontologies ?

Ontologies

- Coming from the field of Ontology (philosophy) :

  study of the nature and structure of reality.

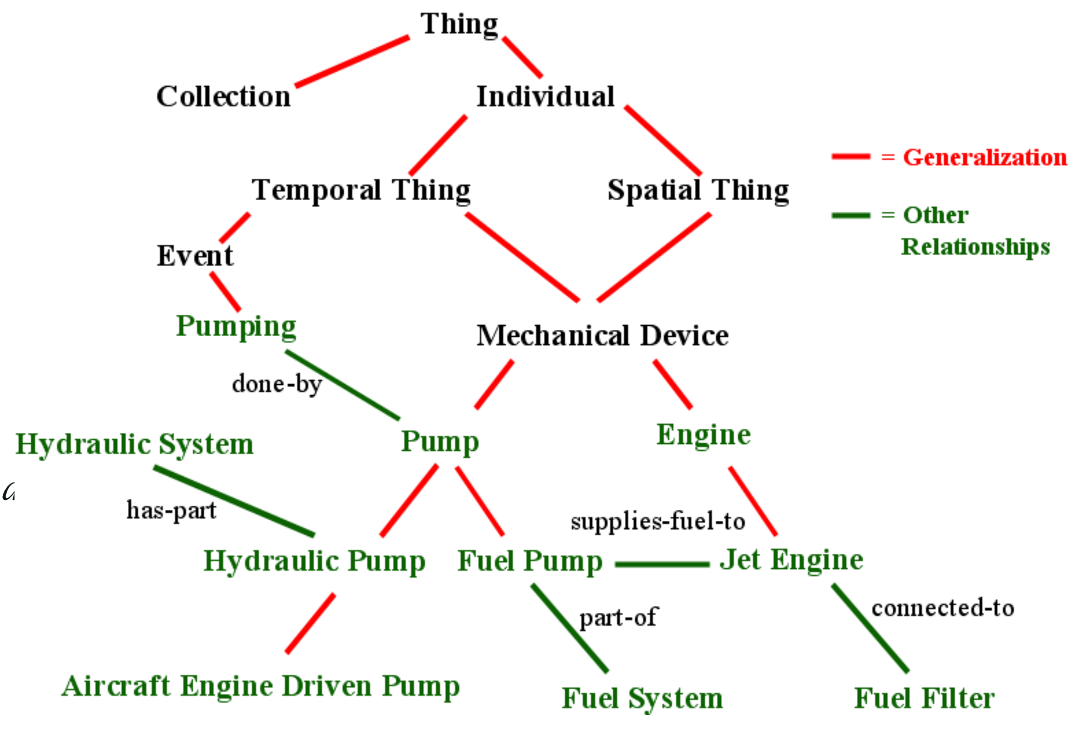
  (Aristotle's *Metaphysic*).

- Computer sciences ontologies (coined by Gruber in 1993) :

  "A set of representational primitives with which to model a domain of knowledge or discourse… typically classes (or sets), attributes (or properties), and relationships (or relations among class members)."
  (Gruber, Encyclopedia of Database Systems, 2009).

  - Introduced in early Artificial Intelligence systems to support inference capabilities, using semantic networks and other knowledge representation schemes.

  - Historically started in Natural Language Processing.



Fragment of an ontology for an aircraft company.

(Formalizated in OWL; source Uschold and Gruniger 2004)

What is common in these situations (3) ?

"A site that exposes RDF usually has an API that is easy to deal with, which makes our life easier. For instance, we use geonames.org as one of our geospatial information sources.

It is a full-on Semantic Web endpoint..."

(interview of Tom Gruber, CTO of Siri, Nova Spivack, 2010).

Semantic Web ?

The semantic web

- Term coined by Tim Berners-Lee to denote a new form of web in which information is given well-defined meaning that can be processed by machines *(Berners-Lee et al., The Semantic Web, 2001)*.

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers… (Berners-Lee and Fischetti, Weaving the web, 1999).

- Not a different web but an extension of the web through various standards used to enrich web data with semantic annotations, organize them and use them.

- These standards are maintained by the World Wide Web Consortium (W3C).
  - RDF (Resource Description Framework).
  - OWL (Web Ontology Language)
  - …

Semantic web and linked data

Share
Add to list
Like
Rate
Tim Berners-Lee at TED2009
Le web à venir par Tim Berners-Lee
16:20

The next web by Tim Berners-Lee (TED talk 2009)*

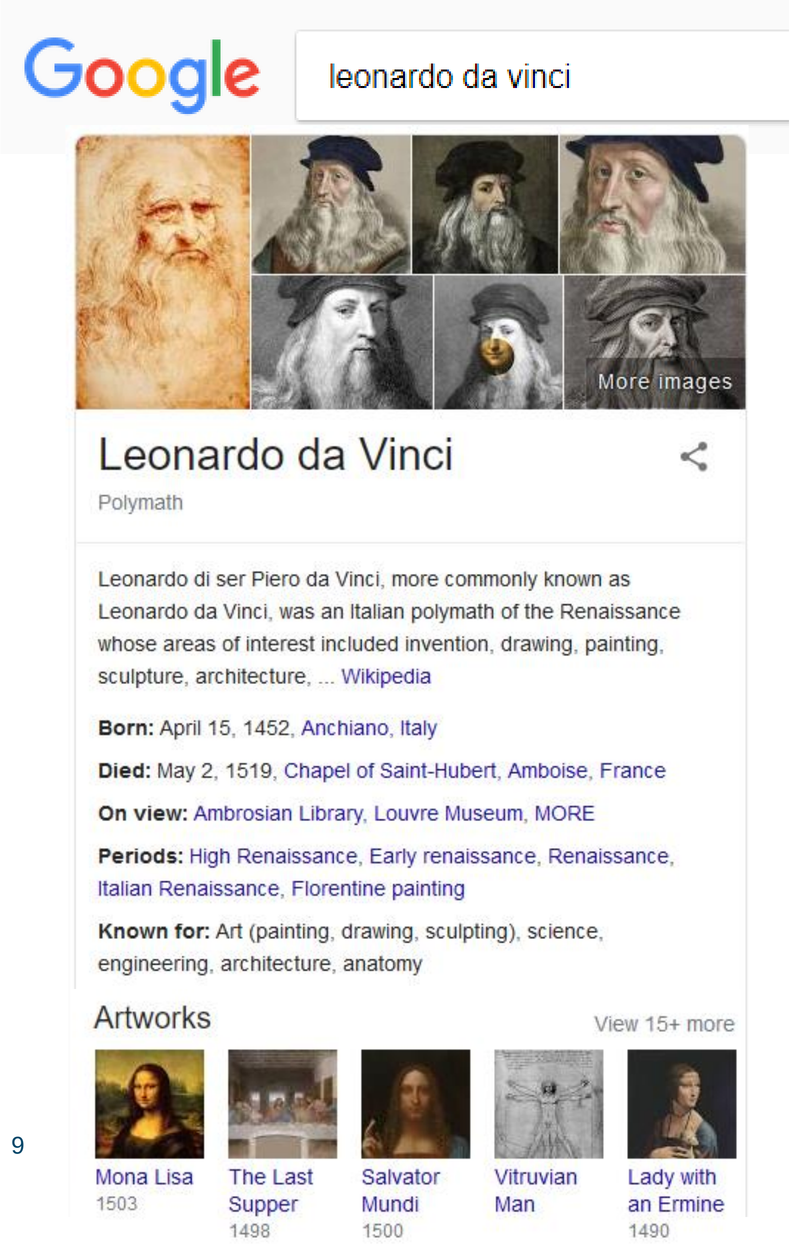
Full talk to be listened to at home.

** : TED is a nonprofit organization started in 1984 devoted to spreading ideas, usually in the form of short, powerful talks (18 minutes or less).*

Key messages :

- ❑ The semantic web will be a web of open linked data !
- ❑ Three rules for linked data :
  - ▪ Names will use the http format.
  - ▪ When reaching these names through the web, we will get back data, in a standard format.
  - ▪ We also will get back relationships. Data is relationships !
- ❑ Linked data is a great way of getting data out of their silos and unlock their power.

What is common in these situations (4) ?



leonardo da vinci

Leonardo da Vinci

Polymath

Leonardo di ser Piero da Vinci, more commonly known as Leonardo da Vinci, was an Italian polymath of the Renaissance whose areas of interest included invention, drawing, painting, sculpture, architecture, ... Wikipedia

Born: April 15, 1452, Anchiano, Italy

Died: May 2, 1519, Chapel of Saint-Hubert, Amboise, France

On view: Ambrosian Library, Louvre Museum, MORE

Periods: High Renaissance, Early renaissance, Renaissance, Italian Renaissance, Florentine painting

Known for: Art (painting, drawing, sculpting), science, engineering, architecture, anatomy

Artworks

View 15+ more

Mona Lisa
1503

The Last Supper
1498

Salvator Mundi
1500

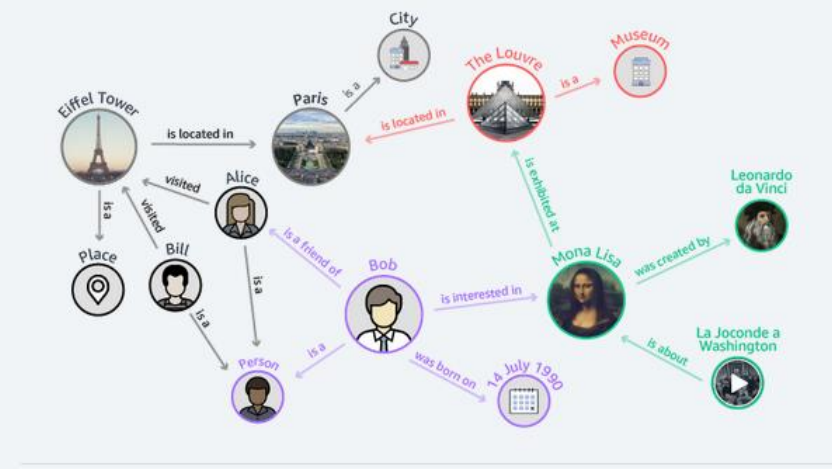
Vitruvian Man

Lady with an Ermine
1490

Introducing the knowledge graph (Google inside blog, 2012)

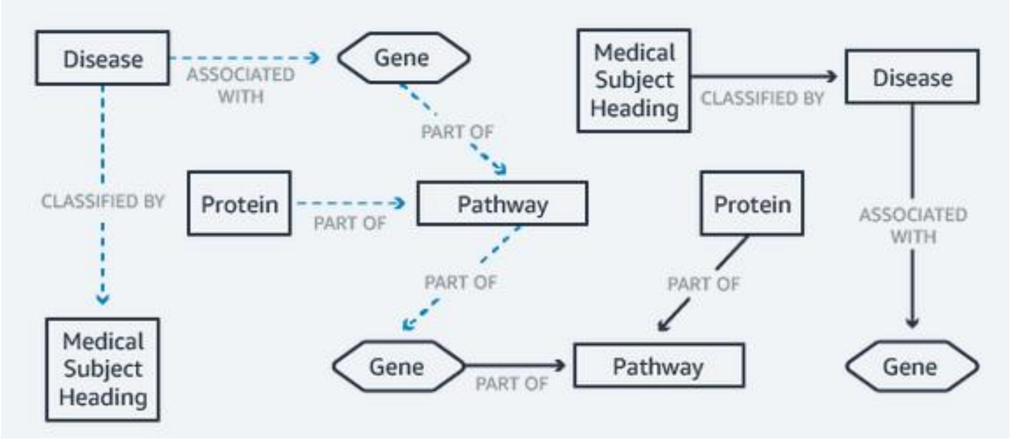
We've been working on an intelligent model— a "graph"—that understands real-world entities and their relationships to one another: things, not strings. (Google inside blog, 2012)

Knowledge graphs



Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities. They have become a powerful asset for search, analytics, recommendations, and data integration (Kroetzsch and Weikum, eds., 2015).

Examples : Google Knowledge Graph, Microsoft Bing, Microsoft Academic knowledge graph, Facebook social graph, eBay product graph, IBM Watson Discovery …



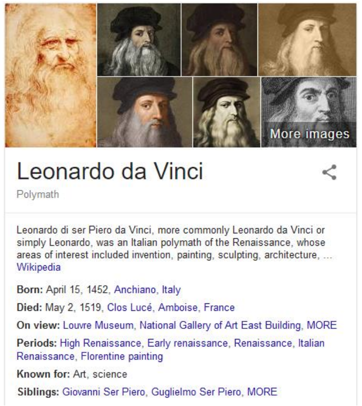
(From Amazon Neptune : new graph database by Amazon Web Services, launched in 2018).

What is common in these situations ?

All use semantic relations and ontologies.

All use the standards of the semantic web.

All make use of meaning-related information : semantic data.



More images

Leonardo da Vinci

Polymath

Leonardo di ser Piero da Vinci, more commonly Leonardo da Vinci or simply Leonardo, was an Italian polymath of the Renaissance, whose areas of interest included invention, painting, sculpting, architecture, ... Wikipedia

Born: April 15, 1452, Anchiano, Italy

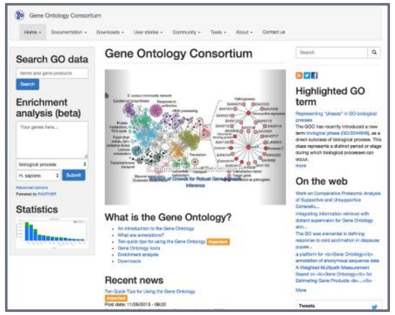
Died: May 2, 1519, Clos Lucé, Amboise, France

On view: Louvre Museum, National Gallery of Art East Building, MORE

Periods: High Renaissance, Early renaissance, Renaissance, Italian Renaissance, Florentine painting

Known for: Art, science

Siblings: Giovanni Ser Piero, Guglielmo Ser Piero, MORE



Gene Ontology Consortium

Home · Documentation · Downloads · User stories · Community · Tools · About · Contact us

Search GO data

Enrichment analysis (beta)

Statistics

Gene Ontology Consortium

What is the Gene Ontology?

- An introduction to the Gene Ontology
- What are annotations?
- Ten quick tips for using the Gene Ontology Important
- Gene Ontology tools
- Enrichment analysis
- Downloads

Recent news

Ten Quick Tips for Using the Gene Ontology Important

Post date: 11/06/2013 - 08:22

Highlighted GO term

Representing "phases" in GO biological process

The GOC has recently introduced a new term biological phase (GO:0044848), as a direct subclass of biological process. This class represents a distinct period or stage during which biological processes can occur.

more

On the web

Work on Comparative Proteomic Analysis of Supportive and Unsupportive Extracellu...

Integrating information retrieval with distant supervision for Gene Ontology ann...

The GO was elemental in defining response to cold acclimation in diapause pupae...

a platform for <b>Gene Ontology</b> annotation of anonymous sequence data

A Weighted Multipath Measurement Based on <b>Gene Ontology</b> for Estimating Gene Products <b>...</b>

More

Tweets

Agenda

1 Introduction to semantic data

2 Outline of the course

Philosophy of the course

- Provide a broad overview of a modern field :
  - Relying on multiple computer sciences disciplines : primarily knowledge representation, but also graphs, complexity, formal languages, databases…
  - Touching different IT domains : Artificial Intelligence, web solutions, big data, software engineering…
  - Multidisciplinary (taking ideas from philosophy, linguistics, logic …).
- With an industrial perspective :
  - From theoretical foundations to business needs.
  - Leading edge solutions in many areas : bio-informatics, engineering, human sciences…
- Concretely applicable :
  - Directly usable standards and languages; practical project with modern tools.

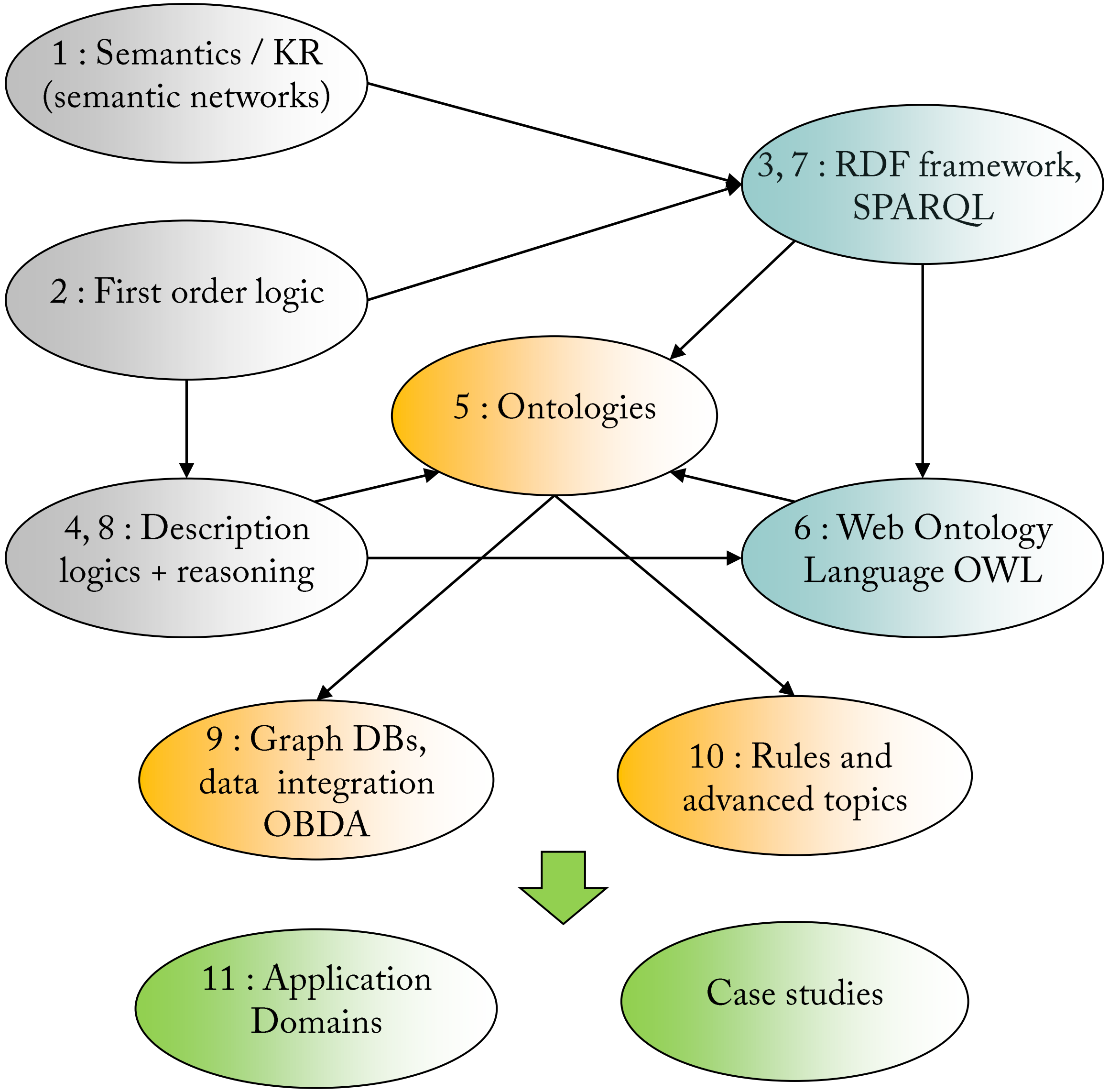
Course content outline

Credits : 5 (theory 25 h, practice 10 h, project 45 h)

Theory (25 h):

1. Semantics and knowledge representation.
2. Introduction to first order logic.
3. The semantic web resource description framework.
4. Description logics.
5. Ontologies and ontology engineering.
6. The Web Ontology Language : OWL.
7. Querying the semantic web : SPARQL.
8. Reasoning with description logics.
9. Data integration and ontology-based data access.
10. Rules and advanced topics.
11. Application domains.

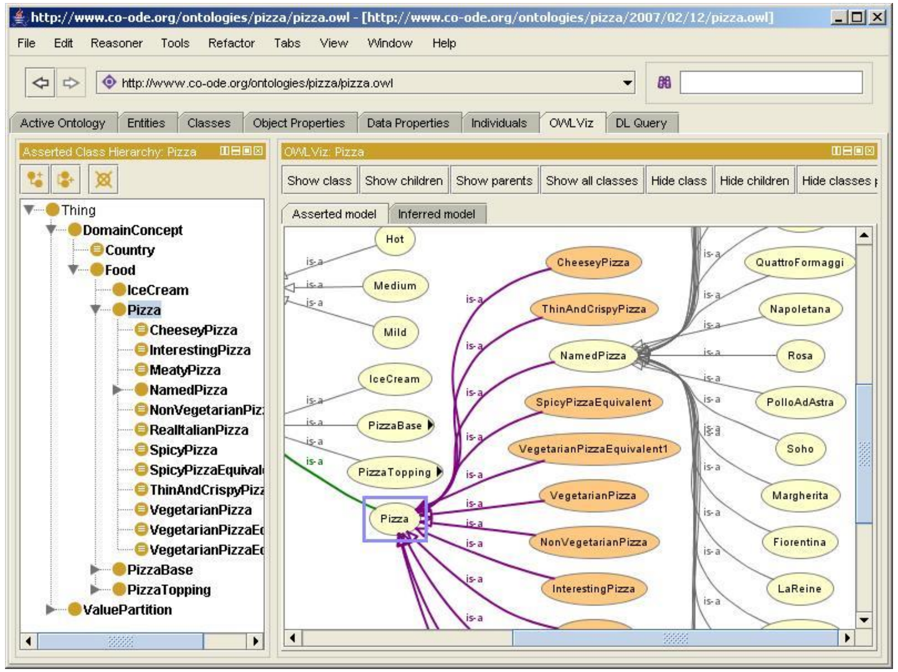
Case studies : real cases for genuine business customers; integrated in the relevant theory sessions.


1 : Semantics / KR (semantic networks)
2 : First order logic
3, 7 : RDF framework, SPARQL
5 : Ontologies
4, 8 : Description logics + reasoning
6 : Web Ontology Language OWL
9 : Graph DBs, data integration OBDA
10 : Rules and advanced topics
11 : Application Domains
Case studies

Course content outline ./.

- **Practice** (10 h) :
  - Exercises on theoretical foundations.
  - Ontology development and tools for the project.

- **Project** (45 h): building and using an ontology.
  - Designing, developing and querying an ontology for a selected domain.
  - Using open-source tools :
    - ontology editor Protégé (Stanford University).
    - Java IDE tool and Java APIs for ontology languages.
  - In groups (3 - 4 people); choice between several topics.
  - Project starting date : March 3, 2021.
  - Topic choices and group compositions will be done before project start.



Organisation

- **Learning material**
  - The course material will be on the course web page : http://www.montefiore.ulg.ac.be/~binot/.
  - The slides of the course provide the reference material for theory and exercises.
  - The domain being multi-disciplinary, there is no single reference textbook. The following sources provide useful but non required input. Additional useful readings are specified in each chapter.

  Artificial Intelligence: A Modern Approach (3rd Edition), Russel and Norvig, 2010, chapters 7, 8, 12.

  An introduction to description logic, Baader, Horrocks, Lutz and Sattler, 2017, chapters 1, 2, 4, 7, 8.

  - The case studies are part of the required course material.
  - For the project, pointers to web documentation of tools and W3C standards will be provided.
- **Submission platform** for the project : Montefiore submission platform.
- **Contact** : Jean-Louis Binot (jean-louis.binot@uliege.be)

Organisation - evaluation

Evaluation modalities may be adapted depending on Covid situation.

- ❑ **Assessment criteria**
  - Theoretical content and related practice : closed books oral exam in June, (open questions and problems).

  All parts of the material are relevant : theory, practice and case studies.

  - Project results will be assessed from :
    - the resulting implementation;
    - a final defense, consisting of a presentation and demonstration, followed by Q&A.
  - Submission of project results is mandatory for presenting the exam !
  - Project results can be kept or improved for September session.
- ❑ **Grade allocation**
  - Oral exam : 60%.
  - Project results : 40% (implementation 25%, defense 15%).

References for this chapter

- *[Baader et al. 2017]: Baader, F., Horrocks, I. Lutz C. and Sattler, U.,* An introduction to Description Logic, *Cambridge University Press, 2017.*
- *[Berners-Lee and Fischetti 1999] : Berners-Lee T. and Fischetti M.,* Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor, *HarperOne, 1999.*
- *[Berners-Lee et al. 2001]: Berners-Lee T. , Hendler J. and Lassila O.,* The Semantic Web, *Scientific American Magazine, May 17, 2001.*
- *[Bodenreider 2008]: Bodenreider O.,* Biomedical ontologies in action: role in knowledge management, data integration and decision support, *Yearbook of medical informatics, 2008.*
- *[Ferruci et al. 2010], Ferrucci D., Brown E., Chu-Carroll J., Fan J., Gondek D., Kalyanpur A., Lally A., Murdock W., Nyberg E.,Prager J., Schlaefer N. and Welty C.,* Building Watson: An Overview of the DeepQA Project, *AI Magazine. 31. 59-79, 2010.*
- *[Gene Ontology Consortium 2012] : Gene Ontology Consortium, "Gene Ontology annotations and resources" Nucleic acids research vol. 41, Database issue, D530-5, 2012.*
- *[Gene Ontology Consortium 2014]: Gene Ontology Consortium, Gene Ontology consortium : going forward, Nucleic acids research* vol. 43,Database issue, D1049-56, 2014.
- *[Google inside blog 2012] : https://search.googleblog.com/2012/05/introducing-knowledge-graph-things-not.html.*
- *[Gruber 2009]: Gruber T.,* Ontology, *entry in the Encyclopedia of Database Systems, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2009.*
- *[Kroetzsch and Weikum, eds, 2015]: Kroetzsch M. and Weikum G., eds,* Journal of Web Semantics Special Issue on Knowledge Graphs, *http://www.websemanticsjournal.org/index.php/ps/announcement/view/19.*
- *[Russel and Norvig 2010]: Russel S. and Norvig, P.:* Artificial Intelligence: A Modern Approach *(3rd Edition), Pearson, 2010.*
- *[Spivack 2010] :* How Siri Works – Interview with Tom Gruber, *CTO of SIRI, blog of Nova Spivack, Jan 16, 2010.*
- *[Ushold and Gruninger 2004]: Ushold M. and Gruninger M.,* Ontologies and Semantics for Seamless Connectivity, *SIGMOD Record, Vol. 33, No. 4, December 2004.*

THANK YOU